ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 95 (2016) 457 – 466
Procedia Computer Science

Complex Adaptive Systems, Publication 6
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2016 - Los Angeles, CA

Shape Analysis of Traffic Flow Curves Using a Hybrid Computational Analysis

W. Kayani[a] S.P. Acharya[a], I.G. Guardiola[*a], D.C. Wunsch[a], B. Schumacher[a], Isaac Wagner-Muns[a]

a. Missouri University of Science and Technology, 600 W. 14 Street. Rolla, MO, USA 65409

Abstract

This paper highlights and validates the use of shape analysis using Mathematical Morphology tools as a means to develop meaningful clustering of historical data. Furthermore, through clustering more appropriate grouping can be accomplished that can result in the better parameterization or estimation of models. This results in more effective prediction model development. Hence, in an effort to highlight this within the research herein, a Back-Propagation Neural Network is used to validate the classification achieved through the employment of MM tools. Specifically, the Granulometric Size Distribution (GSD) is used to achieve clustering of daily traffic flow patterns based solely on their shape. To ascertain the significance of shape in traffic analysis, a comparative classification analysis of original data and GSD transformed data is carried out. The results demonstrate the significance of functional shape in traffic analysis. In addition, the results validate the need for clustering prior to prediction. It is determined that a span of two through four years of traffic data is found sufficient for training to produce satisfactory BPNN performance.

Keywords: Traffic flow; Clustering; Functional data; Granulometric Size Distribution (GSD).

* Corresponding author. *Tel. +1-573-341-6153*
Email address: guardiolai@mst.edu

doi:10.1016/j.procs.2016.09.320

1. Introduction

The ability to accurately analyze traffic flows in an operational setting has been identified as a critical need for Intelligent Transportation Systems (ITS). Previous attempts to accurately analyze traffic flows e.g. volume forecasting models have been restricted mainly to non-functional approaches and methodologies. Daily traffic profiles display functional characteristics (unimodal and bi-modal curves) and are more appropriate for functional analysis rather than traditional non-functional approaches. One of the major advantages of functional analysis is that each daily traffic profile is considered as a single datum, which makes it possible to predict on a much longer term or larger horizon with reasonable accuracy. In the traffic classification domain, previous studies have used non-functional classification to analyze traffic patterns [15], [20] and [4].

The historical literature shows that mainly three approaches are used for traffic analysis: neural networks (NNs); neighbor nonparametric regression and autoregressive integrated moving average (ARIMA) time series models. However, out of these tools neural network is found to be a convenient tool for developing relationships between streams of input and output data, not only for pattern recognition to which they are usually associated, but also for a wide range of modeling situations [12]. Kirby et al. (1997) used a BPNN to carry out a comparative study on NN and statistical models. Neural Networks have performed better than contemporary statistical techniques like discriminant analysis, negative binomial regression, stepwise logistic regression and other classical techniques used in incident detection methodological development [10], [11], gap acceptance modeling [13] and safety modeling [8, 2, 16]. Comparing NN with discriminant analysis highlighted that modeling with NNs places no requirements for a specifying a functional relationship and/or indicates their ability to deal with missing data[8].

As the functional relationship within a neural network is non-linear, it can model poorly defined intricate nonlinear surfaces comprehensively and better in comparison to many traditional linear statistical models. NNs can effectively analyze the patterns from historical data. Other statistical and mathematical models, although proficient in calculation, are often not effective in predictive analysis as they can not adapt to the irregular varying patterns that cannot be easily written in form of a function. In the field of pattern recognition, NNs classify the patterns from training data and recognize if the testing data holds the pattern of interest [14]. In addition, NNs are more responsive to dynamic conditions and do not experience the lag and over-prediction characteristics of time-series models. Owing to the nature of the task at hand, the NN's are considered an appropriate tool for the analysis described herein.

This paper conducts a comparative classification analysis of original historical traffic flow data with and without clustering along with Granulometric Size Distribution (GSD) transformed data using a back propagation neural network (BPNN) as a baseline. Clustering was done using the Partition Around Medoids (PAM) method with a Gap Statistic as the significance test to optimize the stability of the clusters. The "Gap" test (Gap statistics) compares the dispersion of clusters generated from the data to that derived from a sample of null hypothesis tests. The null hypothesis test sets are uniformly randomly distributed data in the box defined by the principal components of the input data. GSD transformed data implies the clustering of unique GSDs generated from each original traffic profile. BPNN is used for training the data's separately and their performance is evaluated by comparing actual and predicted testing targets.

2. Mathematical Morphology

Mathematical Morphology is a theory that provides a number of useful tools for image analysis, which is based on the assumption that images consists of structures that can be handled by set theory. In addition, the tools and methods that MM provides can be naturally applied to the analysis of signals. In traffic the use of image techniques have been mostly applied to tracking of vehicles or determining traffic flow from images of the roadway[1]. In this paper the use of MM techniques are used to realize a unique analysis that lends itself to traditional forecasting of traffic flows. In Gaston et al. 2011, MM tools were applied to signals in an effort to analyze daily solar radiation time series curves[5]. Similarly, in Guardiola et al. (2013) unintended electromagnetic emissions from wireless communication devices are analyzed in the frequency domain[7]. In both previously mentioned works, curves are considered to be bi-dimensional images on which morphological operators are applied to gain information regarding the inherent shape of the two. Thus, a time-series curve is considered as a bi-dimensional image. In this study three MM operators are used to study the shape of the daily traffic flow profile. Specifically, Dilation, Erosion and Opening operators are employed. These operators are used to construct a daily traffic profile's *Granulometric Size Distribution* (GSD).

2.1. Opening, Erosion, and Dilation Operators

In this paper the MM operators are used to study the shape of the daily traffic flow profile. A daily traffic profile is transformed into its representative granulometric size distribution (GSD) function. Consider a time series traffic flow profile represented by function f(t) where it takes only positive values, $f(t) \geq 0$. The MM operators extract the shape of the structure by probing it by a known shape called structuring element (SE). Theoretically, a subgraph of the original function f(t) is defined as $SG(f(t)) = \{(t, y) : 0 \leq y \leq f(t)\}$.

1: Erosion of a function by a SE B[−1, 1] is the function $T_B\,[SG(f(t))] = SG(f(t))\ \theta\ B = \{f(t + b) \mid fb \subseteq B\}$.

2: The dilation of a function by a SE B[−1, 1] is defined as a function $\mu B[SG(f(t))] = SG(f(t)) \oplus B = \{f(t + b) | fb \subseteq B\}$.

3: The combination of the erosion and dilation operations 1 and 2 is called opening. The erosion firstly shrinks the image followed by dilation which expands it. Combining erosion and dilation together to create the opening operation can be expressed mathematically as $\sigma_B[SG(f(t))] = \mu B\{T_B[SG(f(t))]\}$.

Through the employment of these operators we can analyze the shape and functional characteristics of the traffic flow curves.

3. Data Details

The study is based on traffic data collected from the I-94 within the Twin Cities Metro area, Minnesota. Data is obtained at station S110 I-94 East Bound/T.H.65 which has 3 loop detectors D497 94/TH65E1, 170 D498 94/TH65E2 and D499 94/TH65E3. I-94 has 3 lanes in each direction and the station provides composite detector data from three detectors in the eastbound direction. The analysis period is from 1st January 2004 to the 31st of December 2013. However, due to nonavailability of data due to insensitive detectors, the entire 2009 year data is excluded from the study. Partial data is available for years 2011, 2012 and 2013. The detectors measured and logged the flow for each of the three lanes at 30 seconds intervals. For this study, data is aggregated over 15 minute intervals (96 points per day). Weijermars et al. (2005) found that 15 minutes data produce better results as the fine grain variations are removed[18]. Unlike past traffic classification studies Rakha et al. (1995), Weijermars et al. (2005), and Chung (2003) where data consist of 75 days, 118 days and 2 years respectively, this study utilized traffic data of approximately 9 years translating into 2,992 days and 287,232 unique traffic flow value observations at the selected location.

4. Design of Experiment

The study is designed to ascertain the significance of shape in traffic analysis. The concept is as follows: if days are better defined by shape and the shapes vary irrespective of the week days than shape based classification and its onward predictive analysis should be better than methods where shape characteristics are not considered. To this end, the initial assumption is made that every day of the week has unique characteristics e.g. every Monday of the year is the same irrespective of the month and it is also true for rest of the weekdays. With this assumption in mind, a simple target of {Mon=1, Tues=2, Wed=3, Thu=4, Fri=5, Sat=6, Sun=7} is created for the entire 2,992 traffic profiles, refer hereafter as the subjective target. The next step is to cluster the original 2,992 traffic flow profiles using PAM with Gap Statistic as the significance test. The resultant 7 clusters become the target representing clustered data and will be referred to as the original target. Lastly, the entire daily traffic flow profiles are converted to their corresponding GSD curves. Clustering is conducted on the GSD curves. The resultant 7 clusters representing unique shapes become the target representing GSD clustered data and hereafter will be referred to as the GSD target. The original 2,992 traffic profiles has three set of targets: subjective target based on initial assumption of every single week day has similar characteristics, original target based on simple clustering results and GSD target obtained by generating the corresponding GSDs of 2,992 traffic profiles followed by clustering. The 366 original traffic profiles are selected as input and the subjective target {1,2,3, … ,7} for 2004 is selected as output. Similarly 2005 data is employed for testing. The trained network is tested for 2005 data and predicted 2005 subjective target values are compared with actual 2005 subjective testing target. The performance is evaluated by percentage of correct classifying target values. Same process is employed for original and GSD cases by training the data's against respective original and GSD targets followed by their subsequent testing. The performance is evaluated similarly by percentage of correct values of predicted target (output) with actual target. To gain a better insight into the classifying ability of the three cases, a sliding year window methodology is adopted. It implies that the window initially uses one year for training and next year as testing. Subsequently two years are used for training and proceeding year as testing. The process continues until all eight years of data are used as training and the last year's data (e.g. 2013) is tested.

4.1. Clustering Results - Five Basic Shapes

The entire daily traffic flow profiles for 2,992 days are converted to their corresponding GSD curves. Clustering is conducted on the GSD curves using the same method, algorithms and measures. Recall that it is the PAM Algorithm using the gap test statistic method, with a dissimilarity measure of silhouette width to determine the most stable clusters. Out of seven groups obtained, Shape 1 represents Sundays and holidays with typical unimodal shape with peak traffic around 2pm. Shape 3 represents Saturdays, second type of unimodal shape with less significant peak and sustain high traffic around noon to 7 pm. Shapes 4 and 5 represent early and mid week working days behaviors with bimodal shapes. Slight differences exist between the morning and evening peaks of these two shapes. The Shape 2 reflects typical Friday behavior. Although groups 4 and 5 are very similar, they differ immediately after the morning traffic peak. Figures 1a, 1b, 1c, 1d, and 1e represents 95% confidence interval plotted on five clustered groups. The red thick line shows the mean and therefore defines the five distinct shapes obtained through shape analysis of entire dataset of daily traffic flow profiles. The last two groups/shapes representing abnormal behavior (insensitive detector or incidents) are not shown, as there is no dominant shape. Results will differ depending on which 60 percent of the input 253 data is selected for training. As the traffic data is non-linear in nature, sigmoid transfer function is considered more appropriate. The mathematical details are not covered herein due to it being well established (refer to[6]).

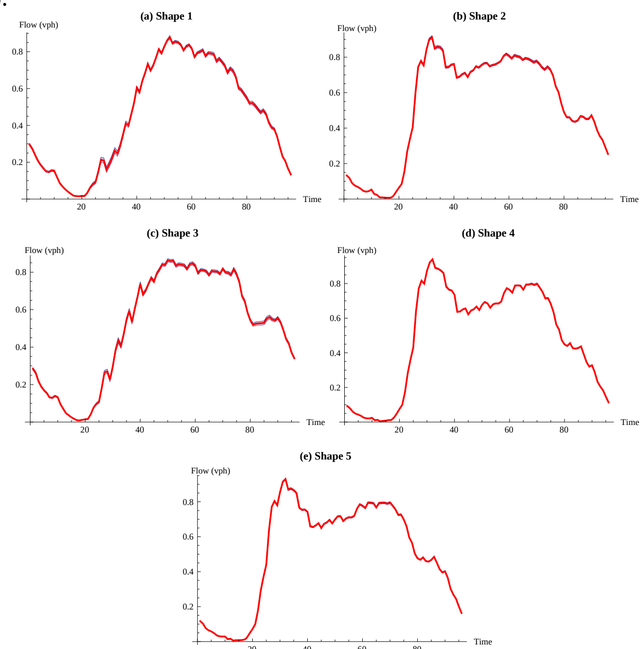

Figure 1: GSD clustering shape based groups.

4.2. Back Propagation Neural Network Methodology

Back propagation training algorithms, when applied to a feed forward multi-layer neural network, is known as Back Prorogation Neural Network. Among back propagation algorithms, Lavenberg Marquardt (LM) is one of the second order methods, which overcomes the slow convergence problem and is widely accepted as most efficient in the sense of realization accuracy [14]. The learning rate is automatically adjusted on each iteration of the algorithm.

During training, the algorithm takes only 60 percent of the input data for training while 20 percent is used for validation and testing respectively. For every attempt of training, the algorithm selects the data randomly from the whole set and not a fixed set of data. Hence, each time the NN is trained results will differ depending on which 60 percent of the input data is selected for training. As the traffic data is non-linear in nature, sigmoid transfer function is considered more appropriate. Mathematical and further details are not covered as the LM algorithm is well established (refer to [6]).

4.3 Network Architecture

Back propagation neural network is used for comparing predictive classification ability of three approaches. A number of studies analyzing traffic data have used a single hidden layer as it produces satisfactory results. One should refer [12] and [17] for examples of applications. In [9], it is concluded that problems that require two hidden layers are rarely encountered, but an NN with two hidden layers can represent functions with a multitude of characteristics. However, there is currently no substantial reason to use NN with more than two hidden layers. In the traffic domain, two hidden layers have been used to achieve better results [21]. The literature does not reveal the best approach in determining the number of neurons within the hidden layers. However, the number of neurons should be determined in such a way that it results in neither under-fitting or over fitting. A typical value for the number of neurons is 30, which is well supported in the literature as a general rule of thumb [12] and [21.] Therefore, a range of 5-30 neurons for single as well as two layer architectures is used to search for the optimal number of neurons. The maximum of correct classifying percentage is taken as the criteria to determine the most appropriate network.

A MATLAB code employing the *newff* function is executed to select the optimal number of neurons using same training dataset. For the single layer architecture, the best performance was found to be one with 12 neurons in the hidden layer with an average correct classifying percentage of 70.375%. In order to get the best two-layer architecture, all possible combinations of neurons in two hidden layers ranging from 5-30 neurons are tested. The best performance is found with the combination of 20 and 5 neurons in two layers. The mean correct classification percentage improved significantly to 76.875%. Therefore, the selected Multilayer Perceptron (MLP) architecture is two hidden layers with 20 and 5 neurons respectively. Table 1 demonstrates the relative performance of the two architectures (e.g. single and double hidden layers MLPs architecture).

Table 1:Details of selected NN architecture

	2005	2006	2007	2008	2010	2011	2012	2013	Mean	
ONE LAYER: Hiden Layers=1, Hidden Neurons=12, Iterations=20										
Correct Class %	77.4	75.9	67.5	67.9	68.4	80	66.5	63.3	70.37	
TWO LAYER: Hidden Layers=2, Hidden Neurons= 20-5, Iterations=20										
Correct Class %	84.2	80.9	77.5	76.4	74.9	81.8	73.7	70.7	76.88	

5. Results and Discussions

5.1 Comparing subjective and GSD output

Firstly, the question of whether every day of the week has a unique shape or days differ in shape owing to the functional and behavioral characteristics is addressed. If the shapes of every day of the week are unique and remain as such, then subjective target should have better classification performance. If the assumption made above is not correct and shape does change, then the GSD output, which is solely based on shape of the traffic profiles, should perform much better than the original output.

A comparison is performed between subjective and GSD output for whole 9 years data with moving year sliding window and 30 training iterations. The input consists of 96 rows and 2,992 columns matrix of traffic data while the output consists of 96 rows of 1 column matrix. The training data is trained and tested initially with the subjective target and later with the GSD target. The result indicates that GSD classification performance is superior to the subjective output. Figures 2a and 2b illustrate the relative classification performance for 2012 and 2013 respectively. It is interesting that GSD maintain a steady classification performance throughout all the years and improved with the increase in the amount of training data. Furthermore, a maximum correct classification of 73% with 8 years training input is achieved. It implies that with more training data, the NN results are improved to a certain level as it gets more training experience with shape profiles and recognizes better. It demonstrates that shape is not a static phenomenon rather its dynamic and varies within days of the week.

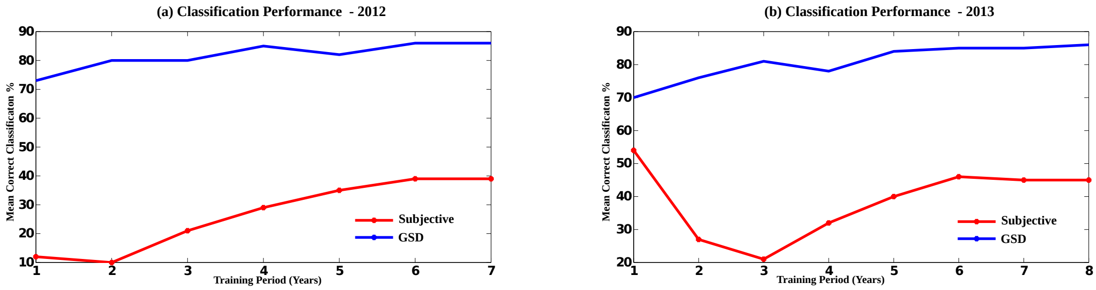

Figure 2: Classification performance of original and GSD.

5.2 Comparison of original and GSD output

In the preceding section the significance of shape is further validated. In this section, the output of the original target trained on original data is compared with the corresponding GSD output trained on GSDs of the same original training data. The comparative analysis is of practical form, as clustering is a common procedure employed when developing traffic prediction or incident detection models. If GSD is better in classification performance than the original output, then it should demonstrate that shape can contribute significantly in improving traffic analysis results and thus suggests that functional approaches are a better option than the prevailing non-functional approaches.

The input consists of 96 rows and 2,992 columns matrix of data for original target and same size matrix of corresponding GSDs for training with GSD target. The training data is trained and tested initially with the original target and later with the GSD target. The results indicate that GSD classification prediction is superior to the original. It shows that days differ in shapes and functional characteristics cannot be ignored in traffic analysis. Figures 3a and 3b demonstrate and illustrate the relative classification prediction performance of original and GSD outputs respectively. In the case of 2012, initially the performance of original target is better than GSD, however with the increase in the amount of available training data the GSD has better classification prediction performance. In the second case, initially both performances are equal but with increase in training data the GSD again gets better results. Another aspect is the initial rise in GSD performance, which becomes stable after being trained for 2 years in 2012 and for 4 years in 2013. It demonstrates that 2-4 years training data is sufficient to train the NN on traffic flow profiles and any further data will not significantly improve the BPNN results.

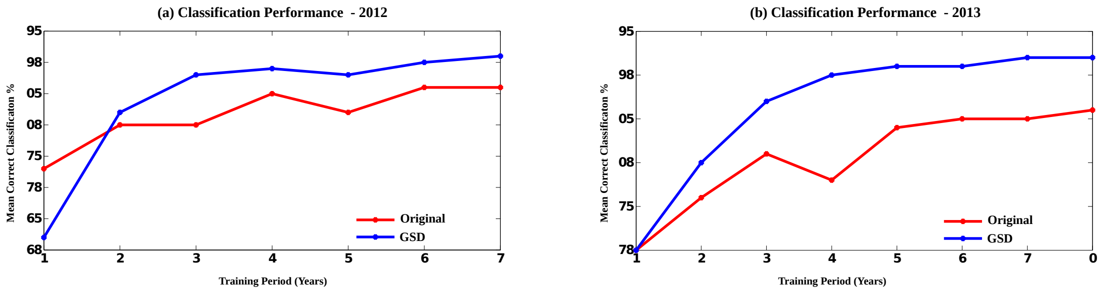

Figure 3: Classification performance of original clustered and GSD clustered.

5.3 Analyzing misclassifications

A confusion matrix is a visual performance assessment of a classification algorithm. To this end, confusion matrices are computed to analyze miss-classified days by the BPNN based on results obtained in section 4.1. Classification prediction of 2013 is carried out based on 8 years worth of training data from (2004-2012) after satisfactory BPNN training. The best outputs of ten training iterations is selected for the original and GSD ensuring that the mean correct classification percentage is in close proximity of mean values already obtained (The resultant two confusion matrices with mean correct classification percentage of 85.43% for original groups represented by `G' and 92.23% for GSD shapes represented by `S'. The total number of misclassifications observed is 24 for GSD and 45 for original. The balanced confusion matrices also suggest that the BPNN architecture is satisfactory. This finding that shape is important is valuable in the analysis of traffic.

Cases of misclassifications are discussed to explain the relative performance of shape analysis over traditional besides explaining the glaring instances of BPNN failure. Figure 4a represent traffic profile of 16 January 2013, which was a working day and defined with conventional morning and evening peaks. However as evident from the figure it has a single insensitive detector reading at interval 68. Owing to this abnormal shape GSD has classified it along abnormal/insensitive detector shapes while the traditional clustering placed it along non working days (Sundays). BPNN classified it along working days not taking into account a major shape deformation. Similarly, the traffic profile of 21 February depicts a working day profile with the exception of dip which might be due to incident or an insensitive detector, The GSD method classified it along abnormal behavior/insensitive profiles, while traditional classified it as non working days (Sunday) and BPNN misclassified it as working day. Figure 4c illustrates the traffic profile for the 7th of March (a working day), however, the shape does not reflect common working day behavior rather represents a non-working day shape along with a dip near the end of the day. Due to shape exhibiting non working day behavior, GSD classified it as non-working day while BPNN predicts it as working day.

Table 2: Confusion Matrix: Original and GSD

Groups	G1	G2	G3	G4	G5	G6	G7	Shapes	S1	S2	S3	S4	S5	S6	S7
G1	9	3	0	0	0	0	0	S1	73	1	0	1	3	1	0
G2	0	0	0	0	0	2	0	S2	0	36	1	0	3	1	0
G3	6	1	64	0	0	7	4	S3	0	0	11	0	0	0	0
G4	0	1	0	50	0	2	0	S4	0	0	0	51	0	1	1
G5	0	0	0	1	45	0	0	S5	5	0	0	1	69	1	1
G6	1	3	5	0	0	75	1	S6	0	0	0	1	0	42	0
G7	3	0	1	0	0	4	21	S7	0	0	2	0	0	0	3

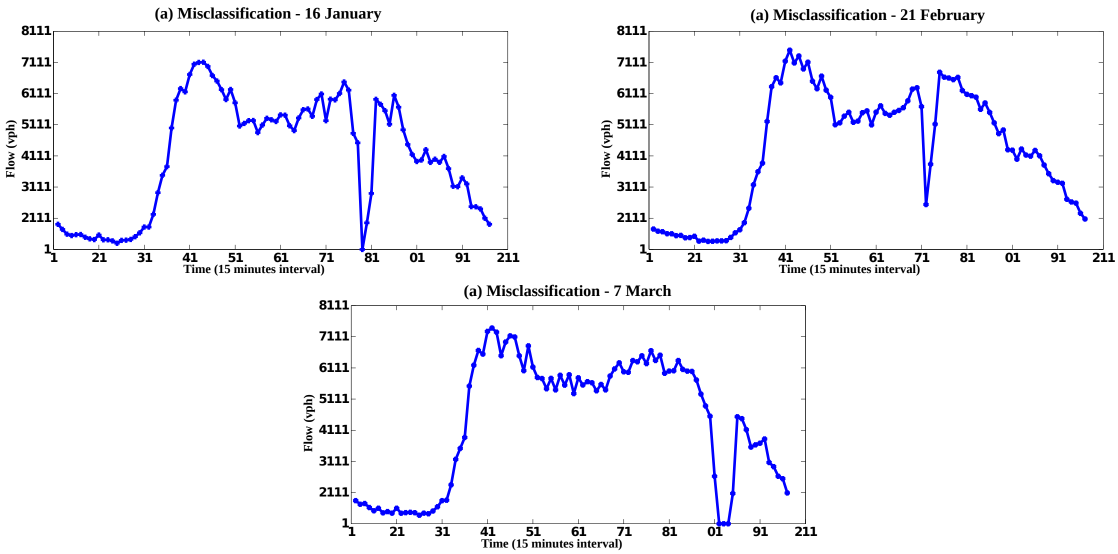

Figure 4: Misclassification examples 2013

The analysis of misclassified cases shows that shape of traffic profiles really change and are not constant for every day. The shape based GSD is quick in identifying an abnormality in shape where traditional methods fail. Although BPNN performed reasonably well in evaluating normal shapes, its performance is questionable in cases where shape exhibits deviated behavior from standard shapes (abnormal shapes due to incident and insensitive detectors).

5.4 Performance of original and GSD targets trained on original traffic profiles

Another comparison is performed by training GSD target on original data rather on GSDs and comparing it with results of original target trained on same data already obtained. The classification prediction performance of clustered is found better in almost all types of yearly training. The performance gap is narrow initially but becomes wider with an increase in training data with the exception of 4 years worth of training data. One understandable reason for GSD having a lower performance is that GSD targets are obtained by clustering GSD profiles and not original profiles. It shows that if shape based MM methodology is to be employed then traffic data has to be dealt in terms of GSDs and not original traffic flow time series curves.

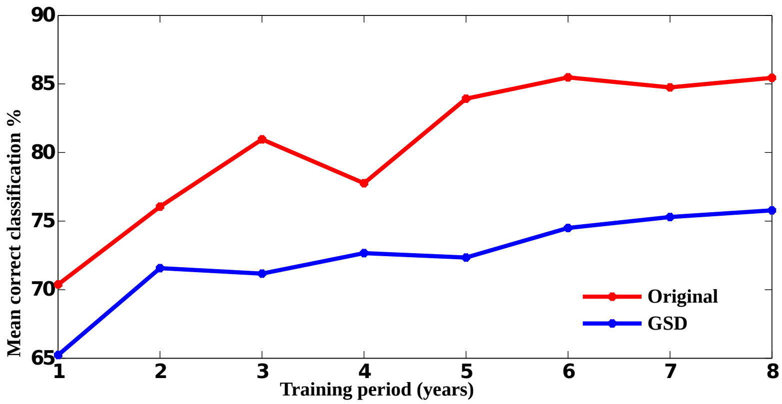

Figure 5: Comparative performance of original and GSD for identical data

5.5 Validating clustering prior to prediction

Clustering explains the hidden structure within the data and provides a simple but meaningful description of data distribution. Irrespective of the prediction algorithm used, the prediction accuracy is negatively affected if data is not fully understood and processed. This study validates that clustering is a necessary step prior to prediction. In previous data with subjective target of arbitrary values {1,2,.....,7} is analyzed. Although no actual clustering is carried out but yet some partition within the data basing on arbitrary values is considered. The predicted output of data with subjective target (arbitrary target values) is far below the predicted output of original target refer figures 6a, and 6b. It is observed that partition of the data and minimizing the distances between the data points with respect to the center point obtained through clustering helps in better classification and prediction. Apart from minimizing the distances within data-set, clustering also contributes towards complexity reduction in the NN due to high similarity all the data and this contribute in enhancing predictive accuracy.

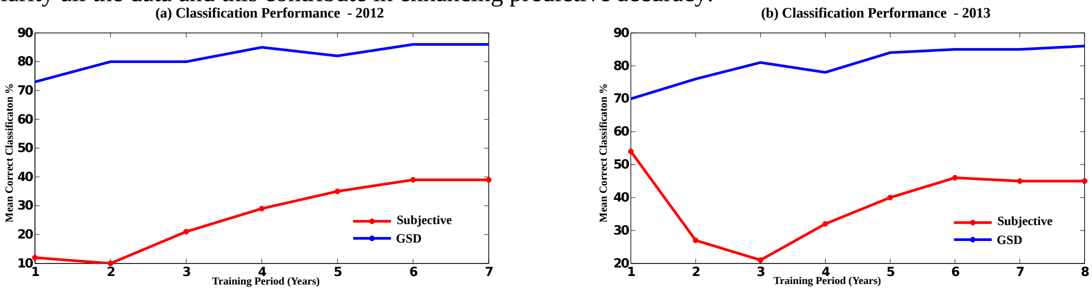

Figure 6: Classification prediction performance of clustering vs. non-clustering

6. Conclusions

This paper concludes that shaped based analysis, clustering and prediction are all substantially increased in performance through the employment of shaped based classification. The study highlights the significance of approaches that analyze complete daily traffic profiles instead of shorter time periods. The comparison between original data (clustered and non clustered) and GSD transformed traffic profiles demonstrate efficacy of shape in classification and prediction. The results show that MM provides a more stable shape based clustering that classifies the existing shape patterns from the training data and recognizes it efficiently during testing.

A major contribution is that shaped based classification methods have the potential to improve the existing traffic prediction models performance by better clustering the volume data prior to the development of a prediction model. Apart from emphasizing the significance of shape, the study also highlighted the necessity of clustering prior to traffic analysis. The performance of BPNN remains satisfactory especially in the context of data used in this study. It is found that 2-4 years of training data is sufficient for training and any further addition does not improve results significantly. MM tools such as the GSD are one of many techniques used in practice for shape analysis. However, other contemporary techniques are also required to be explored. Investigating the functional model to predict traffic profiles is a likely future extension of this study.

Acknowledgements

The authors would like to acknowledge the guidance and assistance in data provision by Minnesota Traffic Observatory, Civil Engineering Dept. University of Minnesota. This work was partially funded through the NSF I/UCRC on Intelligent Maintenance Systems Center at the Missouri University of Science and Technology Award number: 0038586.

References

1. Bai Li, Tompkinson W, and Yan W. Computer vision techniques for traffic flow computation. *Pattern analysis and applications* 2004; 7(4): 365-372.
2. Chang LY, Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. *Safety Science* 2005; 43: 541--557.
3. Chen H, and Grant-Muller S. Use of sequential learning for short-term traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* 2001; 9: 319--336.
4. Chung E. Classification of traffic Pattern. *Proceedings of the 11th world Congress on ITS* (2003): 687--694.
5. Gaston-Romeo M, Leon T, Mallor F, and Ramirez-Santigosa L. A morphological clustering method for daily solar radiation curve. *Solar Energy*, 2011; 85(9): 1824--1836.
6. Goh ATC, Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 1995; 9(3):143--151.
7. Guardiola IG, and Mallor F. A nonparametric method for detecting unintended electromagnetic emissions. *IEEE Transactions on Electromagnetic Compatibility* 2013; 55(1): 58--65.
8. Hashemi RR, LeBanc LA, Rucks CT, and Shearry A. A neural network for transportation safety modeling. *Expert Systems with Applications* 1995; (3):247--256.
9. Heaton J. Introduction to neural network for Jav. 2nd Edition, Heaton Research, 2008; ISBN 1604390085.
10. Ivan JN, and Sethi V. Data fusion of fixed detection and probe vehicle data for incident detection. *Computer-aided Civil and Infrastructure Engineering,* 1998; 13(5):329--337.
11. Khan, Sarosh I., and Stephen G. Ritchie. Statistical and neural classifiers to detect traffic operational problems on urban arterials. *Transportation Research Part C: Emerging Technologies* 1998; 6(5): 291--314.
12. Kirby, H. R., Waston, S. M. and Dougherty, M. S. Should we use neural networks or statistical models for short-term motorway traffic forecasting? *Int. J. Forecasting* 1997; 13: 43--50.
13. Pant, P.D., Balakrishnan, P. Neural network for gap acceptance at stop-controlled intersections. *Journal of Transportation Engineering.* 1994; 120 (3): 432--446.
14. Patra, Prashanta Ku and Vipsita, Swati and Mohapatra, Subasish and Dash, Sanjit Ku. A novel approach for pattern recognition. *International Journal of Computer Applications*. 2010; 9(8) 19--23.
15. Rakha, H. and Van Aerde, M. Statistical analysis of day-to-day variations in real time flow data. *Transportation Research Record*. 1995: 26--34.
16. Sommer C, and Falko D. Progressing toward realistic mobility models in VANET simulations. *Communications Magazine*. 2008; 48(11)132--137.
17. Wei, CH and Ying L. Sequential forecast of incident duration using Artificial Neural Network models. *Accident Analysis and Prevention* 2007; 39(5): 944--954.
18. Weijermars WAM, and Van Berkum EC. Analysis of highway flow patterns using cluster analysis. *Intelligent Transportation Systems*, 2005; 308--313.
20. Wild D. Short-term forecasting based on a transformation and classification of traffic volume time series. I*nternational Journal of Forecasting* 1997; 13(1): 63--72.
21. Yasdi R. Prediction of road traffic using a neural network approach. *Neural computing and applications*. 1999; 8(2): 135--142.